Posted by MiFoo on Sat, 17 Apr 2021 01:27:14 GMT
View Forum Message <> Reply to Message

Hi everyone,

I am using the DHS from 2011 to assess diabetes prevalence and BMI. However, I found a very high number of missing BMI values. Based on the final report (p244/245), there should be 3812 women + 3721 men=7533 people with nonmissing/valid BMI. In contrast to this, I obtained only 5216.7 valid measurements. The numbers per BMI category also deviate, particularly for women. Note that I restricted the analysis to those with valid blood glucose measurements as this is case in the table of the final report.

This is my code (R), with sdata being the survey design object:

```
sdata11 %>%  filter(sh284a<994 &shbm<9998) %>%
summarize(total=survey_total(!is.na(shbm)))

sdata11 %>% filter(sh284a<994) %>%
  mutate(BMI_cat=case_when(shbm<1850~"underweight",
              shbm<2500~ "normal",
              shbm<3000~ "overweight",
              shbm<9998 ~ "obese")) %>%
 mutate(BMI_cat=factor(BMI_cat, levels=c("underweight", "normal", "overweight", "obese"))) %>%

       group_by(BMI_cat, sex) %>% summarize(n=survey_total())
```

	BMI_cat	sex	n	n_se
	<fct>	<fct>	<dbl>	<dbl>
1	underweight	men	1050.	37.0
2	underweight	women	582.	26.3
3	normal	men	2234.	48.0
4	normal	women	808.	30.5
5	overweight	men	303.	18.6
6	overweight	women	169.	15.1
7	obese	men	24.8	5.20
8	obese	women	45.7	7.76
9	NA	men	108.	14.7
10	NA	women	2218.	49.9

Height and weight are available for more people but not their combinations, so that a new calculation of the BMI leads to the same number of missing values and is not the reason for this deviation. I also did the same calculations for the DHS from 2017. For this survey, the numbers more or less match those in the final report. Therefore, I think my coding should not be the explanation for the missing values.

What can be the reason for this? Am I using a wrong variable?

Best wishes,
MiFoo